Experimental Design

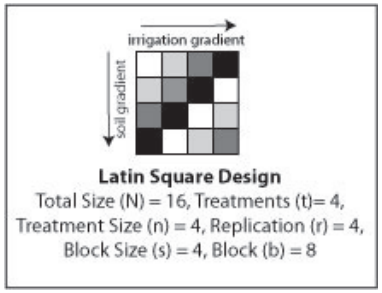
Unit 4

Werner G. Müller

Institute of Applied Statistics

March 26th 2025

5.1 Introduction to Latin squares and related designs (LSD)



With Randomized Complete Block Designs, and the other orthogonally blocked designs it is assumed that the single classification variable associated with blocks accounts for a substantial proportion of the unit-to-unit variation.

Row-Column Designs are used in settings where units can reasonably be sorted by two characteristics rather than one, and the most commonly used of these are *Latin Square Designs* (LSD).

Example: agricultural experiment in a square field, divided into „rows“ and „columns“ of smaller squares. Treatments can be different strains of corn. Units could absorb different amounts of rain water depending on north-south position and different wind conditions depending on their east-west position.

There are two potential sources of nuisance variation, so the unit-to-unit relationships cannot be described independently within rows ignoring columns, or independently within columns ignoring rows.

5.1.1 Row-Column Designs

In order to ensure treatment-block balance comparable to that found in CBDs, it would seem minimally necessary that treatments should be associated with units in such a way that:

- the design is a CBD with respect to rows as blocks, ignoring columns,
- the design is a CBD with respect to columns as blocks, ignoring rows.

So LSD are constructed as per

- the number of row-blocks of units must be t, since each treatment must appear exactly once in each column-block.
- the number of column-blocks of units must be t, since each treatment must appear exactly once in each row-block.
- therefore, a LSD must contain a total of $n = t^2$ units, t of which must be assigned to each treatment.
- each treatment appears once in each row and once in each column.

Row-Column Designs (cont.)

A Latin square is said to be *reduced* (also, *normalized* or *in standard form*) if both its first row and its first column are in their natural order.

Unique Latin squares are often represented in their standard form, equivalent or *isotopic* Latin squares are obtained through reordering of rows, reordering of columns, and/or permutation of the symbols.

Each entry of a $t \times t$ Latin square can be written as a triple (r, c, s), where r is the row, c is the column, and s is the symbol (treatment). A LSD then is a set of t^2 triples called the *orthogonal array representation* of the square.

If we systematically reorder the three items in each triple, another orthogonal array (and, thus, another Latin square) is obtained. E.g. replace each triple (r, c, s) by (c, r, s) which corresponds to transposing the square. The 6 possible reorderings of (r, c, s) give us 6 Latin squares called the *conjugates*.

The number of different Latin squares grows exceedingly quickly with t.

Row-Column Designs (cont.)

Since each experimental unit is contained in two blocks, randomization is somewhat less straightforward for LSDs than with CBDs. For a selected pattern of a LSD, randomization can be accomplished by:

- randomly shuffling the rows of the Latin square, so that each of the $t!$ row orderings is equally likely,
- randomly shuffling the columns of the Latin square, so that each of the $t!$ column orderings is equally likely, and
- randomly shuffling the association of symbols to treatments, so that each of the $t!$ assignments is equally likely.
- the starting standard-form LSD should be randomly selected from the collection of unique Latin squares of the desired size

5.2 Replicated Latin squares

Latin squares can be adjusted in size by adding additional replicates of the entire basic design, i.e. by combining r basic (unique) Latin squares in a design calling for a total of $r \cdot t^2$ units.

These replicates can be thought of as „superblocks“, each containing all the experimental material for a single Latin square.

Where a replicated LSD is used, randomization should be performed independently for each replicate in the experiment.

5.3 A model

A LSD recognizes the possibility of three systematic sources of variation in the data related to rows, columns, and treatments. So a three-way main effects ANOVA model can be used to describe the structure of the data

$$y_{ijk} = \alpha + \beta_i + \gamma_j + \tau_k + \varepsilon_{ijk} \qquad i,j,k = 1,\ldots,t$$

$$\varepsilon_{ijk} \text{ iid with } \quad \mathsf{E}(\varepsilon_{ijk}) = 0 \quad \text{and} \quad \mathsf{Var}(\varepsilon_{ijk}) = \sigma^2 \tag{5.1}$$

where y_{ijk} is the data value observed for the unit appearing in the ith row-block and jth column-block. k is included in the indexing system to identify the effect of the treatment assigned to that unit. Not all possible combinations of i, j, and k are represented in any specific Latin square arrangement.

This additive model has $3t - 2$ degrees of freedom (df), i.e. free parameters to be estimated. With t^2 observations we may always find unique estimates.

A model (cont.)

Two-way interaction terms cannot be meaningfully accounted for in a Latin square design (too many parameters to be estimated on the basis of just t^2 observations).

A model containing an intercept α along with main effects and two-way interactions for row and column (no treatment effect!) would have $\mathsf{df} = t^2$ (free parameters to estimate) - it is *saturated* - it represents all variation in any data set of this form. The residuals would all be zero and an additional treatment effect cannot improve the fit of this model.

Moreover introducing any additional effect to a saturated model yields ambiguous parameter estimates.

In a LSD, inferences can only be made about treatment effects under a model in which the contributions of treatments, rows, and columns are assumed to be additive (no interactions).

Models for replicated Latin squares

For replicated LSD an augmented model is needed to represent the additional sources of variation. If the replicate squares are (physically) unrelated, we may use the model

$$y_{ijkl} = \alpha + \rho_l + \beta_{i(l)} + \gamma_{j(l)} + \tau_k + \varepsilon_{ijkl} \qquad i,j,k = 1,\ldots,t \quad l = 1,\ldots,r$$

$$\varepsilon_{ijkl} \text{ iid with } \quad \mathsf{E}(\varepsilon_{ijkl}) = 0 \quad \text{and} \quad \mathsf{Var}(\varepsilon_{ijkl}) = \sigma^2 \tag{5.2}$$

where ρ_l is the effect due to replicate l. The effects of row-blocks $\beta_{i(l)}$ and column-blocks $\gamma_{j(l)}$ are nested within replicate l, i.e. we have different β_i and γ_j for each replicate l.

Treatment effects are not nested within replicates because they are assumed to be the same in each replicate.

This model has $\mathsf{df} = (t-1)(2r+1) + r$, i.e. free parameters to be estimated. With $t^2 r$ observations we may always find unique estimates.

Models for replicated Latin squares (cont.)

In model (5.2) we assume that row-blocks and column-blocks may have effects. If we assume that the effects of column-blocks are the same in each replicate we get

$$y_{ijkl} = \alpha + \rho_l + \beta_{i(l)} + \gamma_j + \tau_k + \varepsilon_{ijkl} \qquad i,j,k = 1,\ldots,t \quad l = 1,\ldots,r$$

$$\varepsilon_{ijkl} \text{ iid with } \quad \mathsf{E}(\varepsilon_{ijkl}) = 0 \quad \text{and} \quad \mathsf{Var}(\varepsilon_{ijkl}) = \sigma^2 \tag{5.3}$$

This model has $\mathsf{df} = (t-1)(r+2) + r$, of course also here we may always find unique estimates with $t^2 r$ observations.

If we further assume equivalence of row-blocks in each replicate, we end up with

$$y_{ijkl} = \alpha + \rho_l + \beta_i + \gamma_j + \tau_k + \varepsilon_{ijkl} \qquad i,j,k = 1,\ldots,t \quad l = 1,\ldots,r$$

$$\varepsilon_{ijkl} \text{ iid with } \quad \mathsf{E}(\varepsilon_{ijkl}) = 0 \quad \text{and} \quad \mathsf{Var}(\varepsilon_{ijkl}) = \sigma^2 \tag{5.4}$$

This model has $\mathsf{df} = (t-1)3 + r$.

Models for replicated Latin squares (cont.)

Note that the progression through models (5.2) to (5.4) actually represents a strengthening of assumptions being made about the variation associated with blocks.

The number of nuisance parameters decreases from model (5.2) to model (5.4). Models with fewer df will provide more power for tests and narrower confidence intervals. The statistical inferences based on these models - when they are an accurate representation of the system being studied - will be superior to those of models with more df.

It is important that the modeling must accurately represent the actual (physical) experimental situation in order to assure a valid data analysis. At the same time parsimonious modelling is important because fewer nuisance parameters lead to more precision and power in the analysis of the data.

5.3.1 Graphical logic

Because the data collected from a Latin square contains variation associated with row-blocks, column-blocks, and treatments, graphical displays of data by treatment should be „adjusted“ to remove both sets of nuisance effects. Parallel boxplots of

$$y^*_{ijk} = y_{ijk} - (\bar{y}_{i..} - \bar{y}) - (\bar{y}_{.j.} - \bar{y}) - \bar{y} \qquad i,j = 1,\ldots,t$$

are appropriate means for comparisons of the responses associated with each treatment. Expectation and variance for the adjusted response are

$$\mathsf{E}(y^*_{ijk}) = \tau_k - \bar{\tau} \qquad \mathsf{Var}(y^*_{ijk}) = \sigma^2 \left(1 - \frac{1}{t}\right)^2$$

with replicated Latin squares adjustment works analogously.

5.4 Matrix formulation

Beginning with model (5.1) for an unreplicated Latin square, and collecting all nuisance parameters in the first model partition and those associated with treatments in the second, we can write

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\boldsymbol{\tau} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}; \sigma^2\mathbf{I})$$

- $\boldsymbol{\beta}$ is the $(2t+1)$-vector of nuisance parameters α, $\beta_1, \ldots, \beta_t$ and $\gamma_1, \ldots, \gamma_t$ (note: only $2t-1$ free parameters, two zero-sum-conditions)
- $\boldsymbol{\tau}$ is the t-vector of treatment parameters ($t-1$ free parameters)
- $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are n-vectors of responses and random errors where $n = t^2$
-

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{1}_t & \mathbf{1}_t & \mathbf{0}_t & \cdots & \mathbf{0}_t & \mathbf{I}_t \\ \mathbf{1}_t & \mathbf{0}_t & \mathbf{1}_t & \cdots & \mathbf{0}_t & \mathbf{I}_t \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{1}_t & \mathbf{0}_t & \mathbf{0}_t & \cdots & \mathbf{1}_t & \mathbf{I}_t \end{pmatrix} \qquad \mathbf{X}_2 = \begin{pmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_t \end{pmatrix}$$

where $\sum_{i=1}^{t} \mathbf{P}_i = \mathbf{J}$

Matrix formulation (cont.)

$\mathbf{P}_i,\ i = 1, \ldots, t$ are $(t \times t)$ permutation matrices containing zero elements at every position except for a single 1 in each row and column.

The requirement that the permutation matrices sum to $\mathbf{J}$ implies that there is exactly one application of each treatment in any column-block.

$\mathbf{X}_1$ corresponds to β, the $(2t+1)$-vector of nuisance parameters which has only $2t-1$ free elements - hence we have $\mathsf{rk}(\mathbf{X}) = 2t-1$. Some algebra shows that

$$\mathbf{X}_1^T\mathbf{X}_2 = \begin{pmatrix} t\,\mathbf{1}_t^T \\ \mathbf{J}_{(2t\times t)} \end{pmatrix} \qquad \text{is the same as} \qquad \mathbf{X}_1^T \begin{pmatrix} \frac{1}{t}\,\mathbf{J}_{(t^2\times t)} \end{pmatrix} = \begin{pmatrix} t\,\mathbf{1}_t^T \\ \mathbf{J}_{(2t\times t)} \end{pmatrix}$$

Using this equivalence, we have

$$\mathbf{H}_1\mathbf{X}_2 = \mathbf{X}_1(\mathbf{X}_1^t\mathbf{X}_1)^-\,\mathbf{X}_1^T\mathbf{X}_2 = \mathbf{H}_1 \begin{pmatrix} \frac{1}{t}\,\mathbf{J}_{(t^2\times t)} \end{pmatrix}$$

Matrix formulation (cont.)

This is the same as $\mathbf{H}_1\mathbf{X}_2$ for a CRD with t units allocated to each treatment, so an unreplicated LSD and a CRD with the same number of units assigned to each treatment jointly satisfy Condition E. So the reduced normal equations for a LSD are of form:

$$\hat{\tau}_k - \bar{\hat{\tau}} = \bar{y}_{..k} - \bar{y} \qquad k = 1, \ldots t$$

The design information matrix and one of its generalized inverses can be written as:

$$\mathcal{I} = t\,\mathbf{I} - \mathbf{J} = t\left(\mathbf{I} - \frac{1}{t}\mathbf{J}\right) \qquad \mathcal{I}^- = \frac{1}{t}\mathbf{I}$$

Matrix formulation (cont.)

For replicated LSDs, these matrix arguments can be extended by adding the necessary columns of indicator variables for each replicate, and additional necessary columns if row-blocks or column-blocks are nested within replicates, to $\mathbf{X}_1$.

Each of the three models (5.2) to (5.4) of replicated LSDs is Condition E-equivalent to a CRD with rt units assigned to each treatment. The reduced normal equations are then

$$\hat{\tau}_k - \bar{\hat{\tau}} = \bar{y}_{..k.} - \bar{y} \qquad k = 1, \ldots t$$

The design information matrix and one of its generalized inverses can be written as:

$$\mathcal{I} = rt \left(\mathbf{I} - \frac{1}{t} \mathbf{J} \right) \qquad \mathcal{I}^- = \frac{1}{rt} \mathbf{I}$$

5.5 Influence of design on quality of inference

Because Latin square designs are balanced and orthogonally blocked, the reduced normal equations for treatment effects take the same form as those for complete block designs. So we may compare LSDs versus CRDs and CBDs with the same number of runs.

The models of LSD, CRD and CBD differ in their model's degree of freedom df (number of free estimable parameters) and their residual sum of squares $\mathsf{SSE} = \sum_i (y_i - \hat{y}_i)^2$ (the sum of squares associated with the error).

The *residual degrees of freedom* $\mathsf{df}_{\mathsf{SSE}}$ are found by subtracting the model degrees of freedom df from the number of runs n: $\mathsf{df}_{\mathsf{SSE}} = n - \mathsf{df}$.
SSE and $\mathsf{df}_{\mathsf{SSE}}$ are used to estimate the model variance and to test the significance of the treatment parameters:

$$\hat{\sigma}^2 = \mathsf{MSE} = \frac{\mathsf{SSE}}{\mathsf{df}_{\mathsf{SSE}}} \qquad \text{and} \qquad \mathsf{F} = \frac{\mathsf{MST}}{\mathsf{MSE}}$$

Influence of design on quality of inference (cont.)

$\mathsf{F} = \frac{\mathsf{MST}}{\mathsf{MSE}}$ is the test statistic for $H_0 : \ \tau_1 = \tau_2 = \cdots = \tau_t$.

Also the mean of the sum of squares associated with the treatment $\mathsf{MST} = \frac{\mathsf{SST}}{\mathsf{df}_{\mathsf{SST}}} = \frac{\mathsf{SST}}{t-1}$ differs from model to model.

We compare the following models:

- a CRD with rt units assigned to each of the t treatments
- a CBD with rt blocks
- the unreplicated LSD (5.1)
- the replicated LSDs (5.2), (5.3) and (5.4)

The number of runs is the same for all models: $n = r\,t^2$
as is the degrees of freedom associated with the treatment: $\mathsf{df}_{\mathsf{SST}} = t - 1$

Influence of design on quality of inference (cont.)

model	$\mathsf{df}_{\mathsf{SSE}}$	SST
CRD	$t(r\,t-1)$	$\sum_{k=1}^{t} r\,t(\bar{y}_{k.}-\bar{y})^2$
CBD	$(r\,t-1)(t-1)$	$\sum_{k=1}^{t} r\,t(\bar{y}_{.k}-\bar{y})^2$
(5.1)	$t(r\,t-3)+2$	$\sum_{k=1}^{t} r\,t(\bar{y}_{..k.}-\bar{y})^2$
(5.2)	$(t-1)[(t-1)r-1]$	
(5.3)	$(t-1)(r\,t-2)$	
(5.4)	$(t-1)(r\,t+r-3)$	

- The Latin square design can be expected to yield better estimation precision than a CRD with rt units assigned to each treatment if

$$\frac{\sigma^2_{\mathrm{LSD}}}{\sigma^2_{\mathrm{CRD}}} < \frac{\mathsf{t}_{1-\frac{\alpha}{2}}(t(rt-1))}{\mathsf{t}_{1-\frac{\alpha}{2}}(\mathsf{df}_{\mathsf{SSE}}(\mathrm{LSD}))}$$

and better estimation precision than a CBD in rt blocks if

$$\frac{\sigma^2_{\mathrm{LSD}}}{\sigma^2_{\mathrm{CBD}}} < \frac{\mathsf{t}_{1-\frac{\alpha}{2}}((t-1)(rt-1))}{\mathsf{t}_{1-\frac{\alpha}{2}}(\mathsf{df}_{\mathsf{SSE}}(\mathrm{LSD}))}$$

Influence of design on quality of inference (cont.)

- The variance of an estimable function (i.e., contrast) of treatment parameters is

$$\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}}) \;=\; \frac{\sigma^2}{r\,t}\sum_{k=1}^{t} c_k^2 \;=\; \frac{\sigma^2\mathbf{c}^T\mathbf{c}}{r\,t}$$

- For a given estimable function $\mathbf{c}^T\boldsymbol{\tau}$ and signal-to-noise ratio ψ, a desired $\boldsymbol{\Psi} = \frac{\mathbf{c}^T\boldsymbol{\tau}}{\sqrt{\mathsf{Var}(\widehat{\mathbf{c}^T\boldsymbol{\tau}})}} = \psi\sqrt{\frac{r\,t}{\mathbf{c}^T\mathbf{c}}}$ can be obtained with

$$r \;\geq\; \frac{\boldsymbol{\Psi}^2}{\psi^2}\,\frac{\mathbf{c}^T\mathbf{c}}{t}$$

Influence of design on quality of inference (cont.)

The form the noncentrality parameter for the F-test of

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_t$$

is identical to that of a CRD and CBD with $r\,t$ units assigned to each treatment or $r\,t$ blocks

$$\lambda = \frac{1}{\sigma^2}\boldsymbol{\tau}^T\mathcal{I}\boldsymbol{\tau} = \sum_{k=1}^{t}\frac{r\,t}{\sigma^2}(\tau_k - \bar{\tau})^2$$

and the power of this test at level α for given values of τ and σ^2 is

$$1-\beta = Pr(W > \mathsf{F}_{1-\alpha}(t-1;\mathsf{df}_{\mathsf{SSE}})) \qquad \text{where} \qquad W \sim \mathsf{F}(t-1;\mathsf{df}_{\mathsf{SSE}};\lambda)$$

5.6 More general constructions: Graeco-Latin squares

Consider two Latin squares of order t over two sets L and G each consisting of t symbols. These two Latin squares are *orthogonal* if, when superimposed, every pair (l, g) from the Cartesian product $L \times G$ occurs exactly once. If the symbols stand for different treatments each treatment symbol in one Latin square is paired with every treatment symbol in the other Latin square in exactly one cell.

Aα	Bγ	Cβ
Bβ	Cα	Aγ
Cγ	Aβ	Bα

Orthogonal Latin squares were studied in detail by **Leonhard Euler**, who took the two sets to be $L = \{A, B, C, \ldots\}$, the first t upper-case letters from the Latin alphabet, and $G = \{\alpha, \beta, \gamma, \ldots\}$, the first t lower-case letters from the Greek alphabet - hence the name Graeco-Latin square.

Graeco-Latin squares exist for all orders $t \geq 3$ except $t = 6$.

More general constructions: Graeco-Latin squares (cont.)

Orthogonal Latin squares
have been known to predate Euler. In 1725 Jacques Ozanam published a puzzle involving playing cards. The problem was to take all aces, kings, queens and jacks from a standard deck of cards, and arrange them in a 4×4 grid such that each row and each column contained all four suits as well as one of each face value.

The Graeco-Latin Square Design (GLSD) is a direct generalization of the LSD. Suppose we have three sources of „nuisance“ variation with which we must deal, rather than the two accounted for by the rows and columns of a Latin square.

GLSD are Condition E-equivalent to a CRD of the same size with units divided equally among treatments. More general, any blocking arrangement for which the units in each block are evenly divided among the treatments is an orthogonally blocked design and satisfies Condition E.

More general constructions: Graeco-Latin squares (cont.)

A GLSD recognizes the possibility of four systematic sources of variation in the data related to rows, columns, and letters associated with the second Latin square pattern and treatments. So we use a four-way main effects ANOVA model

$$y_{ijkl} = \alpha + \beta_i + \gamma_j + \delta_k + \tau_l + \varepsilon_{ijkl} \qquad i,j,k,l = 1,\ldots,t$$

$$\varepsilon_{ijkl} \text{ iid with } \quad \mathsf{E}(\varepsilon_{ijkl}) = 0 \quad \text{and} \quad \mathsf{Var}(\varepsilon_{ijkl}) = \sigma^2 \tag{5.5}$$

The model has $\mathsf{df} = 4t - 3$ free estimable parameters and t^2 observations, so unique parameter estimates exist only for $t \geq 3$.

All blocks in the GLSD are of size t, and contain each treatment exactly once. It is an orthogonally blocked experimental design, and the reduced normal equations and design information matrix for treatments take the same form as those for CRDs, CBDs, and LSDs in the same number of units for each treatment.

More general constructions: Graeco-Latin squares (cont.)

The form of the full model determines the precision with which the variance of ε can be estimated.

The above model has $\mathsf{df} = 4t - 3$ model degrees of freedom and t^2 observations, hence the residual degrees of freedom are $\mathsf{df}_{\mathsf{SSE}} = (t-1)(t-3)$. This number of degrees of freedom is available for estimating σ^2.

A GLSD can be randomized by randomly selecting a pair of orthogonal Latin squares in normalized form, independently randomizing each of the two LSDs.

6.1 Introduction to Some data analysis for CRDs and orthogonally blocked designs

We are studying the impact of the design on the precision of estimators and power of hypothesis tests in the context of linear models. Since the properties of analysis are the foundation for our motivation to study experimental design, we should discuss the ideas upon which some of these analytical techniques are based.

The quality of the results of our analysis depends largely upon how the assumptions of the linear model are satisfied.

The classical linear model and the generalized linear model (GLM) are extensively studied in the lectures „Verallgemeinerte Lineare Modelle“ from our Bachelor studies and „Advanced Regression Analysis“ from our Master's program. Here we present just a very brief summary of a few widely-applicable techniques. For details we refer to the above lectures.

6.2 Diagnostics
6.2.1 Residuals

Many of the model assumptions may easily be checked with residual plots that indicate whether the residuals have the appearance of an i.i.d. sample from a normal distribution of mean zero and unknown variance.

The standardized and studentized residuals help to circumvent the problem of heteroscedastic residuals. However there remains the problem of correlation among the residuals.

The analysis methods described in the remainder of this section are useful for detecting inequality of variance in CRDs, and interaction between blocks and treatments in CBDs.

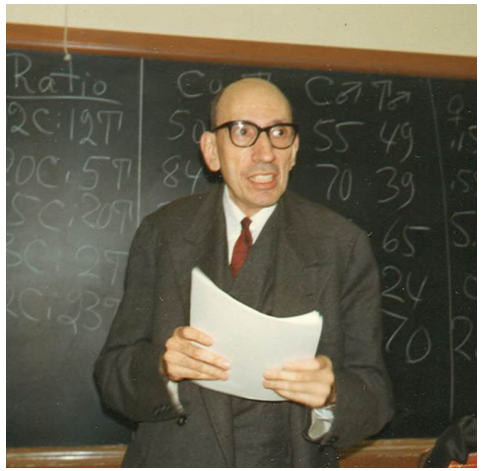
6.2.2 Modified Levene test

The modified Levene test
for equality of group variances, introduced by Levene (1960) is very simple and is performed as follows:

- For each group
  $i = 1, \ldots, t$, compute the median of data values, $\tilde{y}_i$
- For
  each data value, compute the absolute difference between y_{ij} and the associated group median:

$$z_{ij} = |y_{ij} - \tilde{y}_i| \qquad i = 1, \ldots, t \; j = 1, \ldots, n_i$$

- Perform an F-test (one-way ANOVA) for
  equality of means, using the transformed data z_{ij}.



Howard Levene
(1914 - 2003)

6.2.3 General test for lack of fit

A key assumption in all blocked experiments we have discussed is that blocks and treatments do not interact. In most of these designs, the information available to check this assumption is limited.

Only designs that have been enlarged to include „true replication“ yield data in which the variation of model residuals can be compared to the variation within groups of runs with common treatment and block, via a formal test for lack of fit.

The test is presented in section *2.7.1 Pure error and lack of fit* in *Unit 01* and is based on the decomposition of the error sum of squares SSE into the Pure Error sum of squares SSPE and the Lack Of Fit sum of squares SSLOF

6.2.4 Tukey one-degree-of-freedom test

Without „true replications“ the general test for lack of fit described above is not available.

One test for interaction, introduced by Tukey (1949), partitions the SSE into one single degree of freedom component associated with one particular kind of possible interaction, and the remaining $(t-1)(b-1)-1$ degree-of-freedom component which is assumed to represent random noise.

- compute an interaction mean square

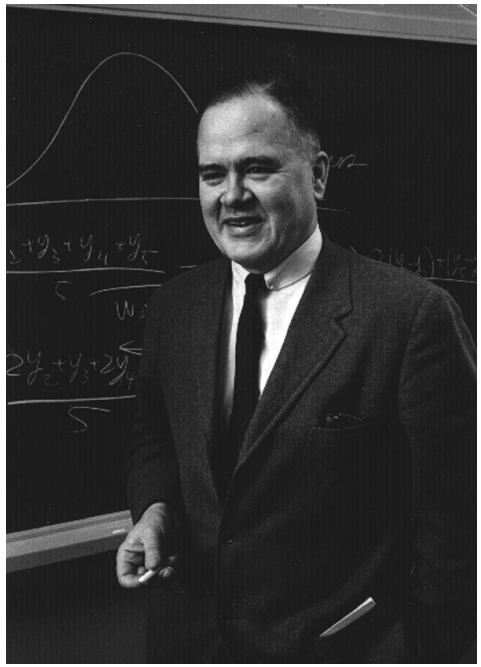
  $$\mathsf{MSI}^* = \frac{\left(\sum_{i,j} y_{ij}(\bar{y}_{i.}-\bar{y})(\bar{y}_{.j}-\bar{y})\right)^2}{\sum_i(\bar{y}_{i.}-\bar{y})^2 \sum_j(\bar{y}_{.j}-\bar{y})^2}$$

  where y_{ij} denotes the observation associated with treatment j in block i.

- and an adjusted error mean square:

  $$\mathsf{MSE}^* = \frac{\mathsf{SSE}-\mathsf{MSI}^*}{(t-1)(b-1)-1}$$

- compare the ratio $\frac{\mathsf{MSI}^*}{\mathsf{MSE}^*}$ to $\mathsf{F}_{1-\alpha}(1;(b-1)(t-1)-1)$ for selected α.



John W. Tukey
(1915 - 2000)

6.3 Power transformations

In many cases where data values are strictly nonnegative, and the variance is inhomogeneous, the variance and mean are related.

A response transformation that equalizes variances over experimental groups, and preserves the desired additive structure for the response mean as a function of treatments and blocks is useful in such situations.

Suppose that our response variable y is actually such that the mean and variance are related through a power law

$$\mathsf{Var}(y) = \mathsf{E}(y)^q$$

Then the approximate variance of y^p variate is

$$\mathsf{Var}(y^p) \approx p^2 \cdot \mathsf{E}(y)^{q+2p-2}$$

Hence, selecting $p = \frac{(2-q)}{2}$ would provide a scale on which the variance of y^p is approximately constant with respect to the mean of y

Box-Cox transformations

Since

$$y^p = \exp(p \ln(y)) = 1 + p \ln(y) + \mathcal{O}((p \ln(y))^2)$$

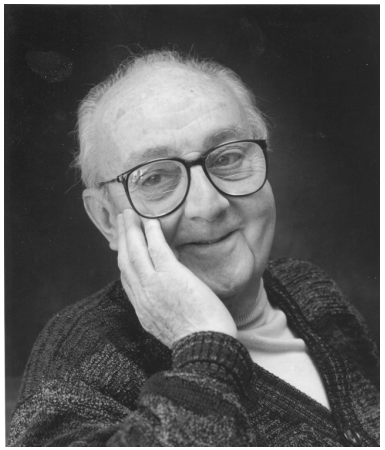
we have

$$y_p^* = \frac{y^p - 1}{p} = \ln(y) + \mathcal{O}(p)$$

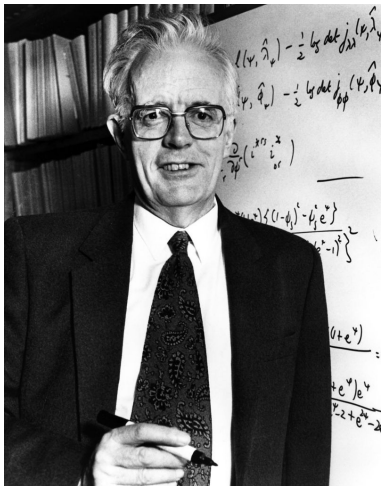
and everything but $\ln(y)$ becomes negligible for p sufficiently small. The above y_p^* is the one-parameter Box-Cox transformation of y.

This power transform is a useful data transformation technique used to stabilize variance and make the data more normal distribution-like.

The only question is how to actually choose the value of p.



George Box
(1919 - 2013)



David Cox
(1924 - 2022)

Box-Cox transformations (cont.)

The optimal value of p can be numerically found as follows:

- Compute the geometric mean of the untransformed data, i.e., $\tilde{y} = \sqrt[n]{\prod_{i=1}^{n} y_i}$
- For a collection of values of p, fit $y_p^{**} = \frac{y_p^*}{\tilde{y}^p}$ to the intended model.
- Choose the value of p that minimizes SSE

In practice, values of p between 0 and 2 are generally of most interest, but many investigators limit attention to $p \in \{0, \frac{1}{2}, 1, 2\}$ unless the data set is large enough to support accurate resolution over a finer grid.

Interpretation of estimates derived under power-transformed data require more attention. We have to use the reverse transformation. The data model implies that

$$\mathsf{E}\left(\frac{y_i^p - 1}{p}\right) = \mathbf{x}_{1i}\boldsymbol{\beta} + \mathbf{x}_{2i}\boldsymbol{\tau}$$

where i indexes the i-th observation. The approximation $\mathsf{E}(\frac{y_i^p - 1}{p}) \approx \frac{\mathsf{E}(y_i)^p - 1}{p}$ leads to

$$\mathsf{E}(y_i) \approx p(\mathbf{x}_{1i}\boldsymbol{\beta} + \mathbf{x}_{2i}\boldsymbol{\tau} + 1)^{\frac{1}{p}}$$

6.4 Basic inference

Up to now the two basic analysis tools discussed in motivating the structure of experimental designs are the F-test for equivalence of treatments, and t-based confidence intervals for specified linear contrasts of treatment effects.

Formulae are simple because CRD is based on one-way ANOVA, and CBDs and LSDs share much of this simplicity of analysis.

In experiments performed to compare a large number of treatments, there may be a need to estimate or test hypotheses about a large number of estimable parameter contrasts.

Whenever we have a multiple testing problem or the dual problem of constructing simultaneous confidence intervals or confidence regions the question of the simultaneous confidence level arises. The way out is the probability of experiment-wise error.

Also this topic is addressed in the lectures „Verallgemeinerte Lineare Modelle“ from our Bachelor studies and „Advanced Regression Analysis“ from our Master's program.

6.5 Multiple comparisons

If we have to test multiple hypotheses the probability for rejecting at least one null hypothesis by mistake is defined as the probability of the experiment-wise error α_E.

If we compute 95% confidence intervals for all 45 pairwise differences of two treatment parameters in a 10-treatment example, the probability that at least one interval for a treatment difference will fail is approximately 0.64. This would be the simultaneous confidence level for the multiple 45 comparisons.

There are several procedures for confidence intervals that maintain a selected experiment-wise type I error probability (Bonferroni method, maximum modulus method, Tukey's method, Fisher's least significant difference (LSD), Scheffé method, Dunnett intervalls etc.).

Here we give only a very brief summary of some methods and refer again to the above lectures.

6.5.1 Tukey intervals

Let T_i be the average of all observations associated with treatment i and $\mathbf{T}$ the t-vector of the T_i.

Tukey uses quantiles from the *studentized range (q) distribution* $q_{1-\alpha_E}(t; \mathsf{df}_{\mathsf{SSE}})$ for the construction of the simultaneous pairwise confidence intervals (α_E is the experimentwise error rate, t is the number of treatments, $\mathsf{df}_{\mathsf{SSE}}$ are the residual degrees of freedom for the model fit, n_i is the number of observations for each treatment)

$$\left[T_i - T_j \pm \frac{1}{\sqrt{2}} q_{1-\alpha_E}(t, \mathsf{df}_{\mathsf{SSE}}) \sqrt{2 \frac{\mathsf{MSE}}{n_i}} \right]$$

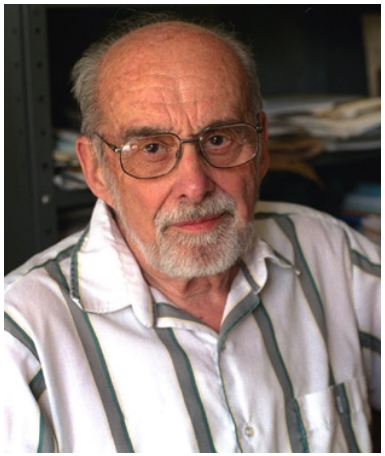
The quantiles for q are tabulated or may be computed with the R-function `qtukey`.

The coverage probability of the above intervals for all possible differences $T_i - T_j$ is at least $1 - \alpha_E$

6.5.2 Dunnett intervals

Dunnett intervals
are a special case of Tukey intervals, they compare every mean to one control mean, e.g. $T_1 - T_i$ for $i = 2, \ldots, t$.

Notice that for these comparisons
the number of parameter estimates is just $t - 1$ which increases the $\mathsf{df}_{\mathsf{SSE}}$ by one. The quantiles of *Dunnett's multivariate* t*-distribution* $d_{1-\alpha_E/2}(t, \mathsf{df}_{\mathsf{SSE}})$ are tabulated. Apart from that the simultaneous pairwise confidence intervals have the same form as Tukey's intervals



Charles Dunnett
(1921 - 2007)

$$\left[T_i - T_j \pm d_{1-\alpha_E/2}(t, \mathsf{df}_{\mathsf{SSE}}) \sqrt{2 \frac{\mathsf{MSE}}{n_i}} \right]$$

The intervals may be computed with the R-function `simint` from the package `multcomp`. Dunnet intervals are smaller than Tukey intervals.

6.5.3 Simulation-based intervals for specific problems

We may generalize Tukey's and Dunnett's idea of simultaneous intervals for a set of c special contrasts $\mathbf{c}_i^T\boldsymbol{\tau}$ for $i = 1, \ldots, c$ where all elements of $\mathbf{c}_i$ are zero except one 1 and one -1.

Let $u_1, \ldots, u_t$ and $v_1, \ldots, v_{df+1}$ be independent random variables following a common normal distribution, say $N(0;1)$, and let $S = \sqrt{\frac{1}{\mathsf{df}}\sum_i (v_i - \bar{v})^2}$ be the sample standard deviation of the second sample. Let further $\mathbf{u}$ be the t-vector of the u_i.

Simulating the distribution of $C = \max_{i=1,\ldots,c}\left(\frac{|\mathbf{c}_i^T\boldsymbol{\tau}|}{S\sqrt{\mathbf{c}_i^T\mathbf{c}_i}}\right)$ gives the quantiles $f_{1-\alpha_E/2}(t, \mathsf{df})$ of the distribution of C.

the simultaneous confidence intervals for all c linear contrasts $\mathbf{c}_i^T\boldsymbol{\tau}$ are

$$\left[\mathbf{c}_i^T\boldsymbol{T} \pm f_{1-\alpha_E/2}(t, \mathsf{df})\sqrt{\mathbf{c}_i^T\mathbf{c}_i\frac{\mathsf{MSE}}{r}}\right]$$

6.5.4 Scheffé intervals

Scheffé intervals are for simultaneous estimation of a collection of **any** contrasts - not just differences.

For any one contrast $\mathbf{c}_i^T \boldsymbol{\tau}$, the Scheffé interval is yet another modification of the t-interval form:

$$\left[\mathbf{c}_i^T \boldsymbol{T} \pm \sqrt{(t-1)\mathsf{F}_{1-\alpha_E}(t-1, \mathsf{df})} \sqrt{\mathbf{c}_i^T \mathbf{c}_i \frac{\mathsf{MSE}}{r}} \right]$$

Henry Scheffé
(1907 - 1977)